http://dx.doi.org/10.21172/ijiet.103.11

An Analysis Of Customer Churn For Mobile Network Operators In Zimbabwe.

Brain Kusotera[1], Fredy Chimire[2], Tichaona W Mapuwei[3]
[1,2,3]Department of Physics and Mathematics, Bindura University of Science Education, Bindura, Mashonaland Central, Zimbabwe

Abstract- The main focus of the study is to analyse factors contributing to churn intension and develop a churn prediction model for mobile network operators in Zimbabwe. A sample of 400 mobile phone subscribers was randomly selected from the customer base of TwoNetwork. Binary logistic regression model, with seven independent variables, was fitted using the data from the MNO. XLSTAT, E-VIEWS and IBMSPSS were used for data analysis and model building. The variables were recharge frequency, call dropouts, off net usage, length of relationship, recharge amount, on net usage and gender. All the variables were found to be significant in influencing the intention to churn. It was found that odds of reducing churn intention were 0.251 times than those of increasing churn per unit increase in recharge frequency. Similarly for length of relationship and on net traffic usage were found to have odds of 5.626 and 1.646 in reducing churn. The odds of reducing churn intension for men were found to be 0.193 times than that of women per unit increase in men. The odds of increasing churn intension were found to be 1.518 times than the odds of reducing churn per unit increase in call dropouts. Similarly, the odds of increasing churn intension were found to be 1.141 times than the odds of reducing churn per unit increase in off net traffic usage. The model developed had a 98% predictive accuracy. TwoNetwork has to enhance customer loyaly through provision of quality services, lower tariffs, always maintain utilities to avoid call drop outs and segment the marget to taraget influential age groups.

Keywords – Customer Churn, Binary Logistic Regression, Mobile Network Operators, OneNetwork (Proxy name for case study company),

I. INTRODUCTION

Intensive competitive pressure is now a global feature in telecommunication market environment and because of this , operators are losing out customers to other operators. According to [9] customer churn refers to a customer leaving a service provider to join another provider.[10], indicated that losing customers has been highlighted in the 21st century as one of the major glitches that vex mobile networks. The annual churn rate ranges from 20% to 40% in most of the global mobile network companies [2].[26] notes that churn rate in developing markets ranges between 20% and 70% with 90% subscribers on prepaid service.Recent studies revealed that current churn rates for mobile wireless telecommunications ranges between 10% and 67%. World annual churn rates vary inversely with growth and profitability [17]. [3] found that a 5% increase in customer retention rates increases profits by 25 to 95%. These reports clearly show the importance of customer churn analysis because the solution of the root causes of the problem, guarantee an increase in profits to the service provider. This means that firms should try to retain their customers to generate customer value. Companies lose at least a quarter of their customers from one year to another. If no new customers are acquired then the average lifetime of an existing customer is equal to 1/c, where c is the annual churn rate. For the company with 25% churn, that means an average lifetime of 4 years for the customer, whereas a churn rate of 50% has 2 years lifetime value [17]. It reported that Churn costs for European and US telecommunications companies estimated around US$4 billion annually while the annual churn rate in telecommunication sector estimated around 30% [22] .Moreover, recently [22] reported that companies globally incur total loss of about US$10 billion annually as a result of switching network provider. Reducing customer churn progressively becomes more important as the telecom markets are saturated.Churn reduction for the company will result in more subscriptions.

In context of recent churn rates in Zimbabwe, [18] (the regulation company) reported that mobile subscriptions for MNOs in the country dropped by 2.5% from 19 477 307 total subscriptions recorded as at 31 December 2015.It indicates that some customers discontinue business relationships with network providers. The table below which display distribution of inactive subscriber base can evidence the existence of churn rates across all networks. In particular, the proportion of inactive customers is for the period of 2016.

Table- 1Quartely churn rates in Zimbabwe

MNO	Total Subcribers	Active Subscribers	%inactive
OneNetwork	9,281,936	6,714,939	27.7%

TwoNetwork	4,278,800	1,824,936	57.3%
ThreeNetwork	5,431,346	4,360,298	19.7%
Total	18,992,082	12,900,173	32.1%

Source: POTRAZ, Operator returns

The table above shows there is serious problem. What then needs to be done to cure this disease? A comprehensive analysis of customer churn is needed coupled with predictive modeling of the intention to churn to equip MNOs. Churn is often treated like flu, it keeps on recurring and great care has to be taken in trying to avoid it [17]. Several studies were conducted on customer churn. In context of studies conducted by [19], a statistically significant correlation between satisfaction and caller abandonment rate attested. The results from his study indicated that a decrease in number of blocked calls on toll free customer service contact could lead to a significant increase in average satisfaction level. The studies have further been supported by [8], where calls blocked ranked third highest in influencing customer satisfaction. However, the study was limited to the relationship between customer satisfaction and call abandoning rate. For instance, it lacks adequate expansion on the model that gives a point estimate of customers who will be unsatisfied because of changes in call abandon rates.
[1], carried out a study on churn management in the mobile market of Sri Lankan mobile market. Binary logistic regression technique was used to find the major causes of churn and revealed that provision of lower tariffs by other network operators as a main contributing factor to customer turnover. [5] did a study on customer churn prediction in the mobile network industry using a case study of MTN Ghana. A data set of 3333 records was used to develop a predictive model with SPSS Clementine. Neural networks and CRT (Classification and regression decision Tree) data mining techniques were applied to determine the propensity of a subscriber to leave a service provider and exhibit of non-churners as well. The predictive accuracy of neural network and decision tree model against the actual churn from training set were 90.37% and 82.56% respectively. The algorithm revealed poor customer service as the most contributing factor to churn.The rest of the paper is organized as follows. The analysis methodology is explained in section II. The results and discussion are explained in section III. Concluding remarks are given in section IV.

II. METHODOLOGY

2.1 Data Collection

A target population with characteristics of interest relating to the research study in pursuit of investigating the causes of subscriber churn and significant churning prediction variables was obtained from secondary database of TwoNetwork Pvt Ltd.The target population comprised loyal and churned TwoNetwork subscribers or subscribers who are about to terminate in the near future. Therefore, target population considers TwoNetwork staff for some departments at kopje plaza and non-staff mobilephone subscribers. Employees of the company are also mobile phone subscribers and have diverse experience on the performance of company network as they stay in different places.
The secondary data was collected from company records of TwoNetwork in the time range of 2010 to 2017 and it enabled the study to accommodate a large sample space that also aids in reliability of findings. Saunders, Phillip and Adrian, advocated that this type of data is readily accessible and minimal time required for collecting [16]. Secondary data have an advantage of it being cheaper to acquire and data collected from a reputable source could be easily verified [16]. However, there exist disadvantages associated with the use of secondary data such as biasedness of information due to errors in the data caused by human manipulation. Despite the shortcomings of secondary data, the researcher used primary data sources to guarantee congruency of research findings. A sample of 400 mobile phone subscribers was randomly selected from the customer base of TwoNetwork. The main causes of customer churn from TwoNetwork were rated on a questionnaire-based survey by respondents of various demographic factors (age, education level, gender, marital status). Systematic random sampling technique was employed inorder to distribute 30 questionnaires. Cronbach's alpha was 0.914 that indicates high internal consistency of the questionnaires and hence the research instrument used for primary data collection was more reliable.

2.2 Churn Analysis and Prediction Model

Churn prediction model is an algorithm used for predicting customer churn. Predicting customer churn is the procedure used for computing the likelihood of future agitating behaviour for subscribers in a database using a predictive model based on prior attributes of churners and non-churners [6]. Churn prediction model aims to identify subscribers, who can easily be convinced to cease the relationship with a company [24]. In another study, it was revealed that logistic regression is a predictive modelling technique that uses historical information on a certain attribute to identify patterns that will assist in predicting future likelihood value [25]. The model is not only used for

prediction but analaysis of the significance of each predictor variable in contributing to churn intention can also be done. This mens that the model provides both churn analysis and prediction of the churn intention as well. Another study pointed out that binary logistic regression is an appropriate model for statistical analysis when the research seeks to assess if a set of explanatory variables predict a dichotomous dependent variable [23]. Binary logistic model is a form of regression, used in a situation when the dependent variable is not continuous. In logistic regression, the estimated value ranges from 0 to 1[20]. Generalized linear models (GLMs) represent a class of regression models that allow us to generalize the linear regression approach to accommodate many types of response variables including count, binary, proportions and positive valued continuous distributions [13-15].Logistic regression is robust to the distribution assumptions of the independent variables.

The main difference between multiple regression and logistic regression is that the value of the dependent variable is estimated in multiple linear regression analysis, while the possibility of occurrence of one of the values which the dependent variable might have is estimated in logistic regression analysis [12]. In this study, it is multivariate binomial logit model since the model has more than two variables. Logistic function transforms variations in the values of the continuous or dichotomous predictor variables on the RHS of the equation to increasing or decreasing probability of the event modelled by the dependent, LHS, variable [4]. The algorithm is chosen because of its better predictive accuracy evidenced from the literature though some variables with an influence in context of MNOs in Zimbabwe were not considered. Based on this algorithm researcher's quest is to understand if explanatory variables stated in the hypothesis can influence and or predict customer churn behaviour. The researcher seeks to ascertain a combination of predictor variables that maximises predictive accuracy of the model in differentiating future churners from the customer database.

2.3 Description of Variables

There are six explanatory variables used by the researcher in the model.The analysis of variables using binomial logit model enable the researcher to come up with significant predictors of churn.

2.4 Recharge frequency(Rf)

Recharge frequency referred as the number of recharge transaction in a month. Recharge frequency for churners used for training in the model we consider recharge transaction frequency up to the date, which subscribers classified as churners. The first variable denoted with x1 in SPSS output.

Call dropouts (Cd).

This measures the number of unsuccessful calls originated by subscribers of TwoNetwork in each quarter because of poor network coverage. In this study, the researcher needs to investigate if a mobile network operator can predict churners based on call failure rates. The second variable also denoted with x2.

Off net traffic usage (Ofn)

Off net traffic usage, measures total minutes consumed by a mobile subscriber for calling other networks.

Length of Relationship (Lor).

This refers to the period of business relationship between a subscriber and a network provider. The variable (Lor) sometimes called relationship longevity.

Recharge amount (Ra).

Recharge amount defined as the total amount of money disbursed by the subscribers in each quarter to meet their communication needs using mobile phone. Recharge amount shall corresponds to the first variable churn prediction algorithm. In relation to churned subscribers, recharge amount is calculated by considering amount spentup to the date they categorized as defectors.

On net traffic usage (On).

On net traffic, usage shall consider MOU for calling subscribers within a network.

Gender (Gn).

This is a categorical independent variable where by male and female are coded with "yes" and "no" respectively. Based on the dummy variable the researcher needs to investigate the sexual group at higher churning risk.

Dependent variable

The dependent variable is dichotomous which is either churner or non-churner. Subscribers who were inactive for at least 30 days are regarded as churners.

2.5 Model specification

Model building starts off with multicollinearity tests. The assumption requires absence of high intercorrelations amongindependent variables. The researcher shall use descriptive analysis inform of correlation matrix among continuous explanatory variables. The assumption of no multicollinearity is satisfied when correlation coefficients

are at most 0.90 [7].The research shall also examine the Variance Inflation Factor (VIF) to assess the presence of multi-collinearity among independent variables. The table below shows the range of possibilities in which VIF will fall.

Table -2 VIF possible ranges

Range	Nature of multicollinearity
0<VIF<5	The is no evidence of multicollinearity problem
0<VIF<5	There is moderate multicollinearity problem
VIF>10	There is a serious multicollinearity problem

After testing for multicollinearity, stationarity test then follows. Although most of the previous research studies reviewed lacked adequate expansion on checking the stationarity of variables used to investigate the problem of customer churn, however the current research find it worthwhile to transcend on investigating the stationarity of times series data inorder to ensure the replicability of the model. A series is said to be stationary if its mean, variance and covariance remain constant over time [21] .The stationarity of data performed on explanatory variables of the churn prediction algorithm except for the categorical independent variable (gender). A unit root test for stationarity called Augmented Dickey-Fuller (ADF) test was employed using Eviews software. Variables are considered stationary if they are I(0) and non stationary if they are I(1+) [21].
The logic functions below represents multivariate binary logit model with seven independent variables.

$$Logit(P) = \beta_0 + \beta_1 X_1 + \cdots + \beta_6 X_6 \quad (1)$$

$$Log(odds) = \beta_0 + \cdots + \beta_k X_k \quad for\ k = 1,2, \ldots \ldots \ldots 6 \quad (2)$$

$$Log\left(\frac{P}{1-P}\right) = \beta_0 + \sum_{k=1}^{k=6} \beta_k X_k \quad (3)$$

Where P is the probability of churning with Y= 1 for all values of Xk, otherwise Y=0 with probability 1-P
The above equivalent formulae symbolize a predictive function of customer churn in this case it is multiple binary logistic regression since independent variables analysed are more than one.
The symbol P represents the probability that response variable Y or customer in a database classified as a churner given k predictors, whereas 1-P is the probability of a customer classified as a non-churner. The study dichotomised the dependent into binary form where churners coded numerically as 1 and 0 otherwise. Logit function is unbounded because it ranges in an open interval of negative and positive infinity. Logit serves as a link function between the probability, P on the RHS and the linear regression expression on the LHS of the binary logistic model. The k variables in the logit algorithm denoted with X corresponds to explanatory variables stated in the research model of the first chapter. Explanatory variables of the research model contains both categorical and continuous data. Independent variables have analogous beta coefficients and sometimes known as parameters of the model which represent the weight of each predictor variable in the logit function.
The probability P predicts the chances with which a randomly selected subscriber in the database has churning risk and is calculated by using a link function of odds or predictors as presented below.

$$p = \frac{odds}{1 + odds} \quad (5)$$

$$p = \frac{\exp(\beta_0 + \cdots + \beta_k X_k)}{1 + \exp(\beta_0 + \cdots + \beta_k X_k)} \quad (6)$$

$$p = \frac{1}{1 + exp - (\beta_0 + \cdots + \beta_k X_k)} \quad (7)$$

for k=1,2,3, ….,8 (8)
Were odds is the ratio of the probability of churning as to the probability of not churn. Data mining was performed by the researchers in to investigate potentially useful variables of secondary information that enable prediction of customers with greater likelihood of ending business relationship with a network provider. The secondary data was used in fitting of logistic regression model to fulfil the objectives of the research. The researcher selected a statistically significant sample of secondary data to come up with a training dataset necessary for model fitting. The sample size computed and data points selected in such a way it provides suretyof adequate representative of customer database and unbiasedness of research results.

III. DISCUSION OF RESULTS

3.1 Multicollinearity tests

The research examined the correlations between variables. It is observed that correlations between independent variables are all at most 0.8. .Therefore, multicollinearity assumption among independent variables met due to absence multicollinearity problem. The only missing variable is the gender, since it is dichotomus.

Table-3 Matrix of Correlates

	Rf	Cd	Ofn	Lor	Ra	On
Rf	1					
Cd	-0.171	1				
Ofn	0.284	-0.500	1			
Lor	0.344	-0.570	0.086	1		
Ra	-0.340	0.746	-0.235	-0.223	1	
On	-0.567	0.171	-0.784	-0.689	0.701	1

** correlation is significant at 0.01 level of significance (two tailed)

The researcher also examined the VIF values to assure of the absence of multicollinearity. From table 4, we can observe that VIF<5 for all predictors suggesting that there is no evidence of a multicollinearity problem.

Table-4 Collinearity statistics

	Collinearity Statistics	
Model	Tolerance	VIF
Rf	0.227	4.396
Cd	0.458	2.185
Ofn	0.236	4.229
Lor	0.224	4.469
Ra	0.331	3.026
On	0.407	2.456

Unit root test

Variables have been tested for stationarity using the ADF and all variables have been found to be stationary in levels at 1% level of significance. Since all variables are I(0) it is possible to estimate the Logit Model.

Table-5 ADF test results

Variable	Probability	Order of integration
Rf	0.0011***	I(0)
Cd	0.0000***	I(0)
Ofn	0.0039***	I(0)
Lor	0.0000***	I(0)
Ra	0.0074***	I(0)
On	0.0007***	I(0)

*** implies stationarity at 1%

3.2 Model Estimation Reults

The classification table indicated that by adding explanatory variables we can now predict categorical variable with overall percentage accuracy of 98.3%, which is an improvement from the baseline model that accounts only for 50% prediction accuracy. Model predicts correct value in 98.3% of the cases.Although the logit predicts churners with greater accuracy there exist AER (Apparent Error Rate), which is the misclassification probability when making churn projections. It can be observed from the confusion matrix that for all 199 predicted churners, three of them are incorrectly predicted.

Table-6 Model estimation results

Variables	B	S.E	Wald	df	sig	Exp(B)	95% C.I for EXP(B)	
							Lower	Upper
Rf	-3.704	0.526	49.658	1	0.022	0.251	0.009	0.069

Cd	0.417	0.122	11.683	1	0.001	1.518	1.195	1.928
Ofn	0.132	0.054	6.022	1	0.014	1.141	1.027	1.268
Lor	-5.626	1.329	17.934	1	0.000	0.004	0.000	0.490
Ra	-0.477	0.163	8.505	1	0.004	0.621	0.451	0.855
On	-1.646	0.796	4.274	1	0.039	0.193	0.040	0.918
Gn	1.172	0.199	34.627	1	0.010	3.230	2.186	4.773
Constant	-1.783	0.670	7.087	1	0.008			

From the table above the odds of reducing churn intension by 3.704 are 0.251 times than the odds of increasing churn per unit increase in recharge frequency. The odds of increasing churn intension by 0.417 are 1.518 times than the odds of reducing churn per unit increase in call dropouts. Similarly, the odds of increasing churn intension by 132 are 1.141 times than the odds of reducing churn per unit increase in off net traffic usage. The odds of reducing churn intension by 5.626 are 0.004 times than the odds of increasing churn per unit increase in Length of relationship. Similarly, the odds of reducing churn intension by 1.646 are 0.193 times than the odds of increasing churn per unit increase in on net traffic usage. The odds of reducing churn intension for men by 1.646 are 0.193 times than that of women per unit increase in men.
By using significant 'B' values, binomial logit coefficients, we can present a churn prediction model as below.
Logit(P)= -1.783-3.704Recharge Frequency+0.417Call dropouts +0.132 Off net traffic usage -5.626Length of Relationship-0.477 Recharge Amount -1.646 On net traffic usage +1.172 Gender (10)

The cut off value is 0.5, that is if p>0.5 a customer is classified as a churner otherwise a customer is loyal (non-churner). All predictors were significant in influencing churn in the customer databank. Binomial logistic regression model predicts churners correctly with 98% accuracy and 98.5% prediction accuracy on non-churners. In comparison with studies carried out by [11],Neural Networks employed yield 84% prediction accuracy of loyal customers where as support vector machine got 84.2 % prediction accuracy of churners. However, the variables included in the current empirical observations incur minimal misclassification costs as compared to the studies carried out by [11]. Therefore the current research findings are grounded on variables that can be extracted from customer data base which adquately fits the churn algorithm.

IV. CONCLUSION

In this paper customer churn analysis was presented for mobile network operators in Zimbabwe using a case study of TwoNetwork Pvt Ltd. Based on research outcome, subscribers can be predicted sufficiently using binary logistic regression with variables namely call drop outs, rechargefrequency, on and off net traffic usage, gender, length of relationship and recharge amount. All the variables were found to be significant in influencing the intention to churn. The model developed had a 90% predictive accuracy. The MNO has to enhance customer loyaly through provision of quality services, lower tariffs, always maintain utilities to avoid call drop outs and segment the marget to taraget influential age groups.

V. REFERENCE

[1] A A. R. E. Adikari, S. U. Rabel, and K. Samarasinghe,"Churn management in Sri Lankan Mobile Market". The institute of Engineers, 37-45, 2008.
[2] A Berson, S. Smith, and K. Thearling, "Building Data Mining Applications for CRM": McGrawHill, New York, 2002.
[3] A Stillwagon, "Small Business Trends", 2014. [Available] https://smallbiztrends.com/2014/09/increase-in-customer-retention-increases-profits.html
[4] AI Schein, "Active Learning For Logistic Regression", 2004.
[5] [available] https://pdfs.semanticscholar.org/771f/fe15acf34e18f9323aedc0b8c12abb741c67.pdf
[6] AA. Kojo," Predicting customer churn in the mobile telecommunication industry", A Cae Study of MTN Ghana, 20111.
[7] A.T. Jahromi, M.M. Sepehri, B. Temourspur and S. Choodbar,"Modeling customer churn in a non-contractual setting: the case of telecommunications service providers", Journal of Strategic Marketing, Volume 18, - Issue 7, 2010.
[8] B. G. Tabachnick and L. S. Fidell,: Using multivariate statistics (6th ed.)". Upper Saddle River, NJ: Pearson Education, 2012.
[9] B.L. Liu "Operationalising Service Quality: Providers' Perspective" Northeast Decision Sciences Institute Proceedings, March 2010.
[10] C. P. Wei," Turning telecommunications to churn prediction: a data mining approach".,Expert Systems with Applications, 103- 112, 2002.
[11] C.F. Tsai and Y.H. Lu,"Customer Churn Prediction by Hybrid Neural Networks", Expert Systems With Applications: An International Journal, Vol. 36, isuue 10, Pp 12547- 12553, 2009.
[12] E. Shaaban, Y. Helmy, A. Khedr and M. Nasr, "A proposed Churn Prediction Model", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622, Vol.2, Issue 2, Pp.693-697, 2012.

[13] H. Bircan," Lojistik regresyon analizi: Tıp verileri üzerine bir uygulama", Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 2, 185-208, 2004.
[14] J .Nelder and R.Wedderburn, "Generalized linear models". J. R. Statist. Soc. A. 135: 370-384, 1972.
[15] J. P. Hoffmann,"Generalized linear models: an applied approach", Pearson: Boston, 2004.
[16] J.Hilbe, "Generalized linear models. American Statistical Association". 48: 255-265, 1994.
[17] M.Saunders, L. Phillip, , & T. Adrian, "Research Methods for Business Students", Italy: Pearson Education Limited, 2009.
[18] N. C. Mupondo, B. Kusotera, D. Mwembe and S.Maposa, "Use of Multistage Optimisation Technique in Formulation of Strategies to Reduce Customer Churn Problem Facing Internet Operators in Zimbabwe", Science Journal of Applied Mathematics and Statistics, Vol. 1, No. 2, pp. 7-24, 2013.
[19] POTRAZ Report 2016
[20] R. K. Feinberg," Operational determinants of caller satisfaction in the contact center",international Journal of Service Industry Management, 131-141, 2002.
[21] R. Tate, " General linear model applications". Unpublished manuscript, Florida State University. 1992
[22] S. Das and T Das, "A Time-series Analysis of Impact of FDI on Economic Development In India during Post-reforms Era (1991-2010)", International Journal of Management, IT and Engineering, Vol.2, Issue 12, ISSN 2249-0558, 2012.
[23] SAS Institute, " Best practice in churn prediction (White Paper)", Cary, NC: SAS Institute, 2000, 2015.
[24] Statistics Solutions,"Data analysis plan: Binary Logistic Regression" [WWW Document]. Retrieved from http://www.statisticssolutions.com/membership-resources/member-profile/data-analysis-plan-templates/binary-logistic-regression/, 2016.
[25] V.L. Migueis, A.S. Camanho, and J.F. Gunha, "Customer attrition in retailing: An application of Multivariate Adaptive Regression Splines", Expert Systems with Applications 40(16):6225–6232, 2013.
[26] V. Sampathi, A.Flagel and C. Figueroa," A Logistic Regression Model To Predict Freshmen Enrollments", Northern Virginia Community College, 2009. [Available] https://analytics.ncsu.edu/sesug/2009/SD016.Sampath.pdf
[27] Wipro Council for Industry Research (WCIR) , "Revenue Enhancement and Churn Prevention for Telecom Service Providers: A Telecom Event Analytics Framework to Enhance Customer Experience and Identify New Revenue Streams" 2013.[Available] http://www.wipro.com/documents/revenue-enhancement-and-churn-prevention-for-telecom-service-providers.pdf